Introduction
Research investigating academic vocabulary has been largely driven by the needs of English as a second or foreign language learners preparing to study in English in higher education contexts. This field of research is growing rapidly along with English for Academic Purposes (EAP). Academic vocabulary sits between “conversational words” and “subject-specific words” in Beck, McKeown, & Kucan’s (2013) well-known three-tiered model of vocabulary. In this chapter, we will look at critical topics relating to academic vocabulary and consider options for future research.

Critical Issues and Topics
This section looks at several key issues and topics relating to academic vocabulary in research, beginning with a short section on why academic vocabulary is important. The next topic looks at definitions of academic vocabulary, followed by the identification of single words as academic vocabulary and then academic multiword units such as collocations, lexical bundles, and formulas in written academic texts. The nature of vocabulary in spoken academic texts is the focus of the following section. Academic spoken vocabulary is a fairly new area of research. It is followed by a short discussion of academic vocabulary in secondary school/middle school contexts. The final topic is academic vocabulary in learning and teaching programs.

Why Is Academic Vocabulary Important?
Research suggests that academic vocabulary is important for several reasons. Firstly, it is a key element of written academic texts. Word list research has found that academic vocabulary can represent anywhere from 10% (Coxhead, 2000) to 14% (Gardner & Davies, 2014) of the words in such texts. These figures suggest that one word in ten or one word in seven in a line of written academic text might be an academic word. Academic study in English-medium institutions requires a large amount of reading, and vocabulary load research suggests that
students need to know a substantial number of words to cope with understanding academic texts. According to Nation (2006), learners need 8,000 to 9,000 word families plus proper nouns to reach 98% coverage of academic written texts. This level of lexical coverage has been found to be sufficient to understand written text (Hu & Nation, 2000; Schmitt, Jiang, & Grabe, 2011). Academic spoken texts also require substantial vocabulary knowledge. Students need to know 4,000 word families plus proper nouns and marginal words to reach 95% coverage of lectures, and up to 8,000 word families plus proper nouns and marginal words to reach 98% coverage (Dang & Webb, 2014). Research suggests that lexical coverage of 95% is sufficient to understand spoken discourse (van Zeeland & Schmitt, 2013). However, increased coverage is likely to increase the number of listeners who can adequately understand speech. Academic word lists are important because they might provide a shortcut to learning the kinds of words that students may often encounter in their academic reading and listening. For example, learning the 1,741 words of the Academic Spoken Word List (Dang, Coxhead, & Webb, 2017) may allow learners to recognize 92% to 93% of the words in academic speech, which is higher than the coverage they may achieve from learning the most frequent 2,000 words of general vocabulary (91%).

Vocabulary testing research suggests that many second language learners of English have low levels of vocabulary knowledge and slow rates of vocabulary growth in a range of contexts, including Denmark (Henriksen & Danelund, 2015) and Taiwan (Webb & Chang, 2012). Learners studying in English-medium institutions need help with vocabulary in general, and they need more help with academic vocabulary in particular because academic vocabulary is found in higher proportions in academic texts than in other kinds of texts, which is an indication of specialization. For example, academic lexis does not occur with the same frequency in general English texts such as fiction (Coxhead, 2000; Gardner & Davies, 2014) and newspapers (Gardner & Davies, 2014) as it does in academic texts. Another important point is that understanding academic texts may be challenging for certain groups of learners. Maxwell (2013) puts this point succinctly by writing, “Nobody is a native speaker of Academic English”. Corson (1995) points out that the social background of learners may have an impact on exposure to academic vocabulary. He coined the term “lexical bar”, and explains that it represents a gulf between the everyday meaning systems and the high status meaning systems created by the introduction of an academic culture of literacy. This is a barrier that everyone has to cross at some stage in their lives, if they are to become “successful candidates” in the “conventional forms of education”.

(Corson, 1995, pp. 180–181)

Coxhead (2000) found that over 80% of her Academic Word List had Greek and Latin roots. It almost goes without saying that learners with first language backgrounds or knowledge of Romance languages will have an advantage over learners who come from languages which do not draw on Graeco-Latin vocabulary. This means learners need to know about word parts in academic vocabulary. See Sasao and Webb (2017) and Sasao (this volume) for more information on word parts and testing this aspect of vocabulary knowledge.

Vocabulary is an important element of university discourse, as Basturkmen and Shackleford (2015) found in their study of first-year accountancy lectures at a university in New Zealand. Explaining and talking about vocabulary was the most frequent language-related episode that occurred among participants in that study. This focus on vocabulary can be seen throughout many areas of teaching and learning in English for Academic Purposes, which demonstrates that there is raised awareness of academic vocabulary in pedagogy as well as
Academic Vocabulary

research. Academic word lists, for example, can be found as integrated sections of English for Academic Purposes (EAP) programs in many places in the world. They can also be found in learner dictionaries, on websites, in learning materials, and vocabulary tests (see Nation, 2016). A practical tool like an academic word list can be immediately useful as a guide for deciding which words to focus on in EAP classes or for independent learning (see Dang, this volume for more on word lists; see also Nation, 2016). Recent research into academic vocabulary research has ranged from large-scale corpus-based studies of written academic vocabulary such as Gardner and Davies (2014) and Browne, Culligan, and Phillips (2013); spoken academic vocabulary (Dang et al., 2017); middle school vocabulary (Greene & Coxhead, 2015); and specialized vocabulary in areas such as discipline-specific vocabulary, for example, business studies (Nelson, n.d.), chemistry (Valipouri & Nassaji, 2013), medicine (Wang, Liang, & Ge, 2008), and chemistry and engineering (Ward, 2009) (see Liu & Lei, this volume for more on technical vocabulary; see also Coxhead, 2018). Academic vocabulary is also a part of corpus-based websites that allow learners and teachers to investigate academic vocabulary in use, for example, the 425-million-word Corpus of Contemporary American English (COCA) (available at http://corpus.byu.edu/coca/) includes academic word lists from which it was derived, and these are now readily available online.

Defining Academic Vocabulary

An important discussion point about academic vocabulary is how it is defined, because definitions drive how it is identified in research. There is a dilemma here (Hyland, 2016), as there are multiple perspectives on academic vocabulary. One way to think about academic vocabulary is that it fits between general vocabulary and technical vocabulary. Indeed, the organization of this volume situates this chapter on academic vocabulary between a chapter on high-, mid-, and low-frequency words and technical vocabulary. This middle ground approach, in turn, identifies academic vocabulary as “items which are widespread in academic discourse, but not very frequent in general English ([for example] establish, evidence)” (Charles & Pecorari, 2016, p. 110). In English for Academic Purposes, this approach is termed “English for General Academic Purposes (EGAP)”. In the EGAP approach, as Hyland (2016, p. 18) points out, “teachers attempt to isolate the skills, language forms, and study activities thought to be common to all disciplines”. Examples of general academic word lists include Coxhead’s (2000) Academic Word List (AWL), Gardner & Davies (2014) Academic Vocabulary List, Browne et al.’s (2013) New Academic Word List, and Dang et al.’s (2017) Spoken Academic Word List. The main idea of these studies was to identify academic vocabulary which all learners in EAP will encounter in their academic texts, no matter what subject area they are studying (Coxhead, 2000; Charles & Pecorari, 2016). A feature of general academic vocabulary, according to Coxhead (2000, p. 124) is that, “Academic words (e.g., substitute, underlie, establish, inherent) are not highly salient in academic texts, as they are supportive of but not central to the topics of the texts in which they occur”.

Gardner and Davies (2014) (see also Nation, 2016) point out that academic vocabulary is actually spread across all frequency levels in English from high- to mid- to low-frequency words. This definition means that words can be academic vocabulary in some contexts and general vocabulary in others, depending on the discipline or context. An example of such vocabulary is area, which is used in everyday English and in mathematics. Some studies have focused on the usefulness of general academic lists in particular subject areas, such as the Academic Word List in agriculture texts (9.06% coverage) (Martínez, Beck, & Panza,
Averil Coxhead

2009) and applied linguistics texts (11.17% coverage) (Vongpumivitch, Huang, & Chang, 2009). For a synthesis of such studies, see Coxhead (2011, 2016).

English for Specific Academic Purposes (ESAP) is based on research into specific domains of study, which Hyland (2016, p. 19) argues is preferable to general academic English because, “In many situations, . . . EAP is most successful when it is tailored to meet the needs of the specific circumstances of the students”. Hyland & Tse (2007) and Dur- rant (2014) argue against general academic word lists on the grounds that the needs of all EAP students cannot be met by one such list. Durrant (2016) has picked up this same issue with Gardner and Davies’ (2014) AVL, drawing on corpora of student writing. See Gardner and Davies (2016) for a response to that article. Hyland and Tse (2007) provide analysis and examples of discipline-specific academic vocabulary meanings and use to make their case against general academic vocabulary. One example is issue, which occurs fairly evenly across science, engineering, and social science with the meaning of topic, but with lower levels of occurrence across the disciplines with the meaning flow out. Wang and Nation (2004) investigated homography in Coxhead’s AWL and found that if the homographs they found in the list were separated, all but three word families would still meet the selection criteria for the AWL. These words are intelligence, panel, and offset. Specialized word lists such as Valipouri and Nassaji (2013) in chemistry, and Ward’s (2009) basic engineering word list, are examples of subject-driven identification of specialized academic vocabulary. Another approach to ESAP vocabulary is to look at a wider discipline such as science, and develop a word list that identifies lexis that occurs across subjects in that area, as Coxhead and Hirsh’s (2007) Science Word List does. This approach adds a layer of specialization for EAP students who might take a range of science-based courses in their first year of university and then choose to continue their studies in a particular area such as biology or chemistry. For more on technical vocabulary, see the chapter by Liu and Lei in this volume.

Nation, Coxhead, Chung, and Quero (2016, p. 150) suggest that the general vs. discipline-specific debate needs a compromise, arguing that “Working on the core meaning and uses of an academic word enables rather than disables current or later learning in more discipline-specific texts”. Hyland (2016, p. 17) also suggests some common ground by envisaging academic vocabulary as being in “a continuum rather than a dichotomy”. Moreover, Dang et al. (2017) point out that both approaches to identifying academic vocabulary practically ignore the fact that EAP learners are not one homogenous group; they have different levels of vocabulary knowledge. This suggests that academic words that are useful for one learner may not necessarily be useful for another.

**Identifying Single Words as Academic Vocabulary**

Early studies focused on identifying academic vocabulary relied on hand counts of lexis in contexts such as student annotations on textbooks, for example Lynn (1973) and Ghadessy (1979). As technology has developed in the last few decades, more recent studies of academic vocabulary have adopted a corpus-based approach. In this section, we will look at different ways to identify academic words in corpora, including common core, corpus comparison, and keyword methodologies, drawing on examples from English for Academic Purposes research. The definitions of academic vocabulary above play out in the identification of academic vocabulary in these studies because whatever definition a researcher adopts will affect the outcome of studies which aim to identify academic vocabulary.

A common core approach is based on the idea that there is a “general pool of language of high frequency items that predominates all uses of language” (Basturkmen, 2006, p. 16).
A feature of the common core approach is an assumption that learners who are focused on academic vocabulary will already have some knowledge of these high-frequency words in English. Coxhead’s (2000) AWL is an example of a common core approach, because it used West’s (1953) General Service List of Words (GSL) to represent high frequency words. This approach did not take into account that some high-frequency words are also academic words (Nation, 2016). As mentioned already, recent research on vocabulary knowledge of second and foreign language learners of English suggests that vocabulary knowledge of high-frequency vocabulary can be patchy, at best (see, for example, Webb & Chang, 2012; Henriksen & Danelund, 2015; Coxhead & Boutorwick, 2018). Basing the AWL, for example, on an existing general service word list means that decisions made for selecting items for West’s (1953) GSL have an impact on the selection of items for the AWL. See Nation (2016), Gardner & Davies (2014), and Hyland & Tse (2007), for example, for critical discussions of the AWL. Browne et al.’s New Academic Word List (2013) is built on a new general service list, developed by the same researchers, and is another example of a common core approach to academic vocabulary.

Other examples of common core approaches to academic vocabulary include Coxhead & Hirsh’s (2007) Science Word List. The focus of this research was lexical items that occurred outside West’s General Service List (GSL) (1953) and Coxhead’s (2000) AWL, and met selection principles from an analysis of a corpus of study guides, laboratory manuals, and textbook chapters from first year university courses in 14 subjects in the sciences.

A corpus comparison approach uses two corpora: a specialized academic corpus and (usually) a general English corpus. This approach allows investigators to identify words which occur more frequently in academic English than in general English. Gardner and Davies (2014) developed the Academic Vocabulary List (AVL) using corpus comparison. This list is based on lemmas rather than word families (see Dang’s chapter, this volume, for more on units of counting in word list development). The academic corpus used by Gardner and Davies (2014) is the 120-million-word academic subsection of the Corpus of Contemporary American English (COCA). This subsection contains nine disciplines, including business and finance; education; humanities; history; law and political science; medicine and health; philosophy; religion, psychology; science and technology; and social science. The academic corpus is made up of journal articles, newspapers, and magazines, and the non-academic corpus contains texts such as magazines and fiction. Table 7.1 shows the top 50 items in the Academic Vocabulary List (Gardner & Davies, 2014).

Nation (2016) highlights a problem with this corpus comparison for academic vocabulary by Gardner and Davies (2014), which is that words such as history, low, and both are included in the list. While the methodology which was used might well select such items, they do not seem to be especially academic in nature.

Identifying academic words by their keyness is also a corpus comparison approach. Keyness studies focus on word frequencies in different corpora, and higher frequency in a specialized corpus is seen as a marker of academic vocabulary. Paquot (2010) developed an Academic Keyword List (AKL) using two academic written corpora (professional writing and student writing by native speakers of English). She drew on principles of keyness, range, and distribution of vocabulary. The AKL includes single and multiword items, and includes high-frequency vocabulary. Some AKL examples are according to, relation to, second, scope, requirement, and late. The AKL can be found at www.uclouvain.be/en-372126.html.

Studies into single academic words illustrate the importance of this lexis. Let’s turn now to academic multiword units.
Identifying Academic Multiword Units

This section begins with the smallest multiword unit, academic collocations (made up of two words), and then moves on to lexical bundles and academic formulas which contain three or more words in strings. Several studies have investigated two-word combinations with a statistical relationship in academic corpora. Coxhead and Byrd (2012) report on patterns in collocations of items from the AWL. They find examples of AWL words occurring in noun phrases which are long and fairly complex, a feature of academic written texts noted by Biber, Johansson, Leech, Conrad, and Finegan (1999; see also Biber, 2006). They also find AWL words that occur together, as in analysis and assessment and analysis and interpretation. Coxhead and Byrd (2012) analyzed the collocations of the AWL word concept and found two main uses. One involved the description and evaluation of a concept (for example, basic, central, dangerous, and different) and the second type referred to the origin of the concept or its specific area of use (as in economic, ideological, legal, and Anglo-American) (p. 13). Another example is assessment, whereby collocates before the word provide information about types of assessment (e.g., brief; appropriate) and collocates after the word

<table>
<thead>
<tr>
<th>Items 1–25</th>
<th>Part of speech</th>
<th>Items 26–50</th>
<th>Part of speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>study</td>
<td>n</td>
<td>history</td>
<td>n</td>
</tr>
<tr>
<td>group</td>
<td>n</td>
<td>develop</td>
<td>v</td>
</tr>
<tr>
<td>system</td>
<td>n</td>
<td>suggest</td>
<td>v</td>
</tr>
<tr>
<td>social</td>
<td>j</td>
<td>economic</td>
<td>j</td>
</tr>
<tr>
<td>provide</td>
<td>v</td>
<td>low</td>
<td>j</td>
</tr>
<tr>
<td>however</td>
<td>r</td>
<td>relationship</td>
<td>n</td>
</tr>
<tr>
<td>research</td>
<td>n</td>
<td>both</td>
<td>r</td>
</tr>
<tr>
<td>level</td>
<td>n</td>
<td>value</td>
<td>n</td>
</tr>
<tr>
<td>result</td>
<td>n</td>
<td>require</td>
<td>v</td>
</tr>
<tr>
<td>include</td>
<td>v</td>
<td>role</td>
<td>n</td>
</tr>
<tr>
<td>important</td>
<td>j</td>
<td>difference</td>
<td>n</td>
</tr>
<tr>
<td>process</td>
<td>n</td>
<td>analysis</td>
<td>n</td>
</tr>
<tr>
<td>use</td>
<td>n</td>
<td>practice</td>
<td>n</td>
</tr>
<tr>
<td>development</td>
<td>n</td>
<td>society</td>
<td>n</td>
</tr>
<tr>
<td>data</td>
<td>n</td>
<td>thus</td>
<td>r</td>
</tr>
<tr>
<td>information</td>
<td>n</td>
<td>control</td>
<td>n</td>
</tr>
<tr>
<td>effect</td>
<td>n</td>
<td>form</td>
<td>n</td>
</tr>
<tr>
<td>change</td>
<td>n</td>
<td>report</td>
<td>v</td>
</tr>
<tr>
<td>table</td>
<td>n</td>
<td>rate</td>
<td>n</td>
</tr>
<tr>
<td>policy</td>
<td>n</td>
<td>significant</td>
<td>j</td>
</tr>
<tr>
<td>university</td>
<td>n</td>
<td>figure</td>
<td>n</td>
</tr>
<tr>
<td>model</td>
<td>n</td>
<td>factor</td>
<td>n</td>
</tr>
<tr>
<td>experience</td>
<td>n</td>
<td>interest</td>
<td>n</td>
</tr>
<tr>
<td>activity</td>
<td>n</td>
<td>culture</td>
<td>n</td>
</tr>
<tr>
<td>human</td>
<td>j</td>
<td>need</td>
<td>n</td>
</tr>
</tbody>
</table>

*Note:* The parts of speech noted in Table 7.1 are n = noun, v = verb, j = adjective, and r = adverb.

*Source:* Gardner and Davies, 2014, p. 317
provide some sort of characterization of the item which is being assessed (e.g., assessment of success; assessment of change) (Byrd & Coxhead, 2012, p. 12).

Durrant’s (2009) study of academic collocations involved a 25-million-word academic written corpus that contained the following disciplines: arts and humanities; engineering; law and education; medicine and health sciences; science; and social sciences. He used a keyword analysis to compare the occurrences of collocations in the academic corpus with a non-academic corpus. Durrant (2009) identified 1,000 collocations, and found that 763 of them were grammatical, in that they contained one non-lexical word (for example, assume that, associated with, and based on). This finding is a reminder of the importance of high-frequency vocabulary in academic texts (actually, in all texts, see Dang et al., 2017), and highlights the problem with the exclusion of high frequency words in Coxhead’s AWL (Durrant, 2009).

Another study of academic collocations resulted in Ackermann and Chen’s Academic Collocation List (ACL) (2013), which was developed using a common core approach by analyzing the Pearson International Corpus of Academic English (PICAE). PICAE contains journal articles and textbooks from 28 disciplines and contains over 25 million running words. PICAE has four disciplines: applied sciences and professions, humanities, social sciences, and natural/formal sciences. The collocations were identified in the corpus using a computer program, and the data was then refined using expert ratings until the final list of 2,468 items was reached (downloadable from http://pearsonpte.com/research/academic-collocation-list/). Ackermann and Chen (2013) categorized their Academic Collocation List according to grammatical patterns. Almost three-quarters of the Academic Collocations List were in a noun combination category (for example, assessment process and classic example).

The studies highlighted here illustrate that corpus-based studies into academic collocations result in a great deal of data, and it is no mean feat to work through the data sets to decide what might be worth investigating further, what meets selection criteria, and what might be the most useful items for teachers and learners. Note that both studies which have been discussed here present large numbers of academic collocations which would be quite daunting to take into English for Academic Purposes classrooms.

Moving beyond two-word combinations takes us to lexical bundles, which are words in a string of three or more words which occur frequently (Biber et al., 1999). Examples of lexical bundles include on the basis of, on the other hand, and at the same time. Biber, Conrad, and Cortes (2004) found that lexical bundles in academic texts occur more often in spoken classroom language than in textbooks or academic written prose. An explanation for this finding is that the lexical bundles in classroom discourse are “useful for instructors who need to organise and structure discourse which is at the same time informational, involved, and produced with real-time production constraints” (Biber, 2006, p. 148). Biber (2006) found more lexical bundles in natural science than in business, engineering, humanities, and social science, which he associates with the heavier technical content of natural science.

Differences in frequency and lexical bundles in disciplines have also been noted by Hyland (2008), Pickering and Byrd (2008), and Byrd and Coxhead (2010). Like Biber, Conrad, and Cortes (2004), Pickering and Byrd (2008) found more lexical bundles in spoken than written academic texts. Byrd and Coxhead (2010) found that arts, commerce, law and science shared 73 four-word lexical bundles. One example of a shared bundle is on the other hand. It had a total frequency of 353 in Coxhead’s academic written corpus from the AWL study, and occurred most often in the law subsection of the corpus (35%), followed by commerce (27%), arts (27%), and then science (15%). Law contained the highest number of lexical bundles at 5.44%, followed by commerce at 2.65%. Arts and science both had lower
levels of lexical bundles at around 1.45%. Hyland (2008) also found low levels of lexical bundles in applied linguistics.

Another study of multiword units containing strings of more than three words was carried out by Simpson-Vlach and Ellis (2010), who were in search of pedagogically useful academic formulas. Simpson-Vlach and Ellis (2010) used a quantitative analysis of corpora to identify academic formulas (the Academic Formulas List – AFL) in academic written English and academic spoken English, drawing on statistical analyses and comparisons with non-academic written and spoken corpora. They then asked experienced language teachers and language testers to rate a sample of formulas based on whether they were a formulaic expression, were cohesive, and worth teaching. Three lists were developed from this process. One list is the core AFL list, which contains both written and spoken formulas (for example, *in terms of*, *at the same time*, *from the point of view*, *in order to*, and *as well as*). The second list contains 200 spoken academic formulas (for example, *be able to*, *blah blah this is the*, *you know what I mean*, and *you can see*), and the third list contains 200 written academic formulas (for example, *even though the*, *a wide range of*, *was based on*, *take into account the*, and *as can be seen*). The formulas were also categorized according to their functions in texts, such as ability and possibility (*allows us to*; *are able to*) and evaluation (*an important role in*; *is consistent with*). Note that the formulas are made up of high frequency words in strings. This point is important when we think about the nature of academic vocabulary and multiword units: high frequency and non-content words predominate in these patterns (Durrant, 2009).

A more wide-ranging study of academic multiword units comes from Liu (2012) who investigated the occurrences of lexical bundles, phrasal verbs, and idioms, drawing from studies such as Biber et al. (1999), Carter and McCarthy (2006), Simpson-Vlach and Ellis (2010), Gardner and Davies (2007), and also from dictionaries. Liu (2012) used the academic sections of the Corpus of Contemporary American English (COCA) and British National Corpus to look for these multiword units in the corpora and he ranked the resulting list of 228 frequent multiword units into three frequency bands. The first band has 77 units which occur 100 times or more in the corpora; the second bank contains 85 units which occur between 50 and 99 times in the corpora, and the final band contains 67 units which occurred between 20 and 49 times. Some examples from the first band include *according to* (*det + N*), *as well as* (*det + N*), and *NP suggest that*.

All of these studies into multiword units in academic written English suggest that there are many such units, in many patterns. Byrd and Coxhead (2007) point out that bringing such patterns into classrooms can be quite tricky, and learners tend to use fewer lexical bundles in their academic writing than professional writers (see Cortes, 2004).

### Academic Vocabulary in Spoken Academic English

Spoken academic vocabulary in English is a much less-researched topic than written academic vocabulary. Until the advent of large scale academic spoken corpora, such as the British Academic Spoken Corpus (BASE) (see Thompson & Nesi, 2001) and the Michigan Corpus of Academic Spoken English (MICASE) (Simpson, Briggs, Ovens, & Swales, 2002), it was much more difficult and expensive to obtain spoken academic texts for analysis. The lectures in the BASE and MICASE corpora have featured in an analysis of lexical bundles in academic speaking (Nesi & Basturkmen, 2006), who focused on the cohesive discourse in the lectures and the use of referential bundles and discourse organizers, such as *at the same time* and *if you want to*.
The vocabulary needed for listening to academic speech was explored in Dang and Webb’s (2014) analysis of the BASE corpus. This study found that the most frequent 4,000-word families plus proper nouns and marginal words provided 95% coverage of the corpus, and the most frequent 8,000-word families plus proper nouns and marginal words provided 98% coverage. Dang and Webb (2014) also noted that learners can reach 95% coverage of academic spoken texts if they know the 3,000 most frequent word families in English and Coxhead’s (2000) AWL. This study suggests that spoken academic texts require fewer general academic words than written academic texts. Dang and Webb (2014) found that the AWL accounted for 4% of spoken academic texts. A study by Thompson (2006) reported that the AWL covered 4.9% of a corpus of lectures. For more on academic vocabulary and listening to lectures, see Rodgers and Webb (2016).

Research has also looked into the vocabulary of listening assessments. Webb and Paribakht (2015) analyzed texts used in a Canadian university admission test called CanTEST and found that the most frequent 4,000-word families covered 95% of these texts but that the most frequent 10,000-word families were needed to reach 98% coverage. In another study, Paribakht and Webb (2016) found variation of AWL coverage of reading and listening passages in a university admission test, with higher coverage in reading (6.31%) than listening.

A more recent study by Dang, Coxhead, and Webb (2017) on vocabulary in academic spoken texts has resulted in an Academic Spoken Word List (ASWL) of 1,741-word families. The ASWL was developed from a corpus of 13 million running words from lectures, seminars, labs, and tutorials. This word list took a core academic vocabulary approach and as well as corpus-based measures, and it included teacher and student-based data. A key feature of the ASWL is it can be adapted according to the vocabulary knowledge of different learners. Dang et al. (2017) found that knowledge of ASWL can help learners reach between 92% and 96% coverage of academic spoken English.

**Academic Vocabulary in Secondary School/Middle School Contexts**

Much research on academic vocabulary has been situated in higher education contexts, and yet it is also important to secondary school or middle school education (Crossman & Pinchbeck, 2012; Greene & Coxhead, 2015; Roessingh, 2016). One study that looks into academic vocabulary in middle school textbooks in the US is based on Jennifer Greene’s 18 million running word middle school corpus of 109 textbooks from grades 6 to 8 in the subjects English grammar and writing, health, mathematics, science, and social sciences and history (Greene & Coxhead, 2015). West’s GSL (1953) accounted for nearly 80% of the words in the corpus, and Coxhead’s AWL covers nearly 5.4%. The middle school vocabulary lists from each of the five subject areas contained items which only occurred outside the GSL. Selection criteria for the middle school lists include frequency and range from within each subject area, leading to words in the science list such as energy, organism, element, and react; and equation, graph, fraction, and data in the mathematics list. Each middle school vocabulary list contains between 600 to 800 word types, and coverage of their respective subject-based corpora range from the lowest at 5.83% in social studies and history through to 10.17% in science (Greene & Coxhead, 2015, p. 23).

Several studies have looked into coverage of Coxhead’s AWL (2000) in secondary school texts. Roessingh (2016) found higher proportions of the AWL in expository texts (9.86%) than literary texts (5.5%). In a study of academic vocabulary in the teacher talk of three teachers (maths, science, and English as an additional language) in an international school in Germany for grade 6 students (aged 10–11 years), Coxhead (2017) found that the coverage
of the AWL on average was 1.92%. This figure is around half the coverage reported by Dang and Webb (2014) of the AWL in university-level spoken texts. The percentage of AWL words in the teacher talk increased in all three subjects over the course of a year, from around 1% in each subject to 2.05% in science, 2.45% in English as an additional language, and 2.78% in mathematics. Coxhead, Stevens, and Tinkle (2010) examined the occurrence of Coxhead’s AWL in a corpus of secondary school science textbooks and found the list covered 7.05%. This coverage was 2% lower than the AWL coverage in the science subcorpus from the original AWL study which was made up of university-level texts. Coxhead et al. (2010) also investigated the coverage of Coxhead and Hirsh’s (2007) Science Word List (317 word families) and found that it covered 5.90% of their corpus of secondary school science textbooks, which is higher than the coverage reported by Coxhead and Hirsh in university-level science texts (3.79%). These figures suggest that the Science Word List is potentially more useful for secondary school learners than students preparing for university studies. For more on coverage of the Science Word List in specialized texts, see Coxhead and Quero (2015).

**Academic Vocabulary in Learning Programs**

The final topic in this section concerns research into academic vocabulary in programs of study. There is a growing interest in this research, because identifying academic vocabulary to support EAP learners and their teachers is one thing, but putting the results of such studies into practice and evaluating the effectiveness of approaches to teaching and learning is quite another. Storch and Tapper (2009) report on increased and appropriate use of AWL words in EAP student writing over time in an Australian context, while Crossman and Pinchbeck (2012) report on increased sophistication in academic vocabulary use in writing by seven Generation 1.5 learners in a university preparation course in Canada. Luxton, Fry, and Coxhead (2017) report on gains in academic vocabulary knowledge over a six-month period. Their study involved 2,642 students in 35 secondary schools in Aotearoa/New Zealand, and was based on responses on the academic section of Schmitt, Schmitt, and Clapham’s (2001) Vocabulary Levels Test. The participants were roughly divided between monolingual speakers of English and bilingual/and or multilingual English-language speakers, and participants studied a range of subjects in secondary school, including English literature, English as a second language, physical education, religious education, technology, and science. Teachers were provided with suggestions for focusing on academic vocabulary development in classes and were supported in their attempts to focus on this lexis. Statistically significant gains in test results were reported by Pasifika students, Māori-medium (kura kaupapa) students, students who spoke languages other than English at home, second language speakers of English who had spent between three to five years in New Zealand, and those who had attended English language support classes. Luxton et al. (2017, p. 14) also found high coverage figures (19%) for Coxhead’s AWL in assessment texts at the secondary school level. The importance of academic vocabulary in assessment texts was highlighted in interviews with secondary school teachers (Coxhead, 2018).

There is value in finding out more about how learners develop their understanding and use of academic multiword units. Jones and Haywood (2004) reported on an attempt to include 80 lexical bundles (such as there were no significant differences and studies have shown that) selected from work by Biber et al. (1999) into EAP coursework, and to measure student use of these multiword units in their writing. Jones and Haywood (2004) found that the classroom activities helped raise the learners’ awareness of the lexical bundles, but this awareness was not accompanied by much evidence of actual use in writing. Another example of such research is
Li and Schmitt’s (2009) study of a Chinese first language MA student’s use of multiword units (lexical phrases) in writing over an academic year in a British university.

**Future Directions**

This chapter has demonstrated that while there is a growing literature on academic vocabulary, the majority of this research has focused on the identification of this lexis. Moreover, these studies have tended to employ a range of corpus-based methodologies and predominantly target undergraduate-level education. More research is needed to examine academic vocabulary in secondary school and postgraduate education in different contexts, and draw on both written and spoken data. Both single-word and multiword unit analyses are needed for these contexts. It is also important that corpus-based research is complemented by qualitative approaches, as can be seen in the work by Simpson-Vlach and Ellis (2010) and Dang et al. (2017). More studies are also needed into academic spoken events other than lectures. There is also a need to find out more about how academic vocabulary is learned and how this vocabulary develops over time.

Another important area for future research is replication studies. Miller and Biber (2015) examine approaches to validation of word lists, including academic word lists, and draw attention to the need to find out more about how different corpora can have an impact on the selection of items and whether the same results would or could be found using different corpora. Replication is also important in academic vocabulary learning and teaching studies.

Another direction for future research is to investigate the extent to which academic vocabulary occurs in languages other than English. A couple of studies are already underway in this area, including, for example, Danish at the University of Copenhagen by Anne Sofie Jakobsen (see Jakobsen, Coxhead, & Henriksen, 2018) and Welsh as part of the CorCenCC (Corpus Cenedlaethol Cymraeg Cyfoes – The National Corpus of Contemporary Welsh) project, led by Dr. Dawn Knight at Cardiff University (go to www.corcencc.org/). The predominance of research in this chapter is based on English for Academic Purposes, but this work needs to be balanced by ground-breaking research in other languages and in different levels of education.

**Further Reading**


This volume contains chapters which relate to or directly discuss aspects of academic vocabulary.


This book contains a chapter on specialized vocabulary and word lists.


This book has chapters on academic vocabulary in secondary school contexts and university in English-medium contexts.

**Related Topics**

Classifying and identifying formulaic language; frequency as a guide for vocabulary usefulness; high-, mid-, and low-frequency words; technical vocabulary; word list; and key issues in researching multiword items.
References


Academic Vocabulary


